植 物 分 类 与 资 源 学 报 2012, 34 (3): 211~221
Plant Diversity and Resources DOI: 10.3724/SP.J.1143.2012.12041

初探低拷贝核基因在低等分类阶元系统发育重建中的适用性——以十字花科为例*

丁 铭, 曾丽萍, 马 红, 张 宁**
(复旦大学生命科学院, 植物科学研究所, 进化生物学中心, 上海 200433)

摘要: 分子系统学已被广泛用于解决物种之间的亲缘关系。迄今为止, 被子植物分子系统学大都采用叶绿体和线粒体基因。但叶绿体和线粒体多数采用单亲遗传方式, 不能完全记录物种的进化历史。而且叶绿体和线粒体基因相对保守, 可用于系统发育重建的信息有限, 很难用来解决科以下类群之间的亲缘关系。相反, 核基因遵守双亲遗传方式并能提供大量信息位点, 但却没有得到广泛运用。本文以十字花科为例, 从17种十字花科植物中得到了5个编码蛋白的单拷贝核基因序列, 采用最大简约法、最大似然法和贝叶斯法重建它们之间的亲缘关系。结果表明, 和目前常用的基因相比, 这些基因能提供更多的信息位点, 由此得到的系统发育树具有很高的支持, 利用最大简约法得到的5个核基因的系统树中各分支均得到了100%的自展支持率。结果表明, 我们选取的核基因可以用来进行科、属及种间的系统发育重建。因此, 这5个基因可以用来解决被子植物其它科内的亲缘关系, 而且也可以作为DNA条形码研究的有效分子标记。

关键字: 分子系统学; 十字花科; 低拷贝核基因; *MCM5*; *MLH1*; *AT3G54630*; *AT3G26730*; *AT2G47990*
中图分类号: Q 949, Q 78 **文献标识码**: A **文章编号**: 2095-0845(2012)03-211-11

The Use of Low-copy Nuclear Genes for Reconstructing the Phylogeny of Low-level Taxonomic Hierarchies: Evidence from Brassicaceae

DING Ming, ZENG Li-Ping, MA Hong, ZHANG Ning**
(*Institute of Plant Biology*, *Center for Evolutionary Biology*, *School of Life Sciences*, *Fudan University*, Shanghai 200433, China)

Abstract: Molecular systematics is widely used to determine organismal relationships. However, until now, systematics of angiosperms has mostly employed organellar genes, but these genes are generally inherited uniparentally, only partially reflecting the evolutionary history. In addition, chloroplast and mitochondrial genes are highly conserved, with limited numbers of phylogenetically informative characters. In contrast, nuclear genes are inherited biparentally and have relatively large numbers of informative sites, but only few nuclear genes have been utilized. In this work, sequences of five low-copy nuclear genes were obtained in 17 species from Brassicaceae and used for phylogenetic analyses with the Maximum-Parsimony, Maximum-Likelihood and Bayesian methods. Highly similar topologies were obtained using the five genes; in addition, compared with organellar genes, the nuclear genes had more phylogenetically informative sites and provided stronger supports. Specifically, the supporting values of all the nodes were 100% in MP trees using the five nuclear genes. Therefore, these five genes are excellent candidate marker genes for plant taxonomists who aim to resolve the relationships among low-level taxonomic hierarchies and might also be useful for DNA barcoding.

Key words: Molecular systematics; Brassicaceae; Low-copy nuclear genes; *MCM5*; *MLH1*; *AT3G54630*; *AT3G26730*; *AT2G47990*

* 基金项目: 国家自然科学基金委资助项目 (31100156); 中国博士后基金资助项目 (20100480549, 201003241)
** 通讯作者: Author for correspondence; E-mail: peaceful_ning@hotmail.com, zhang_ning@fudan.edu.cn
收稿日期: 2012-03-26, 2012-05-04 接受发表
作者简介: 丁 铭 (1985-) 男, 硕士研究生, 主要从事十字花科系统发育研究工作。E-mail: 09210700153@fudan.edu.cn

物种间的系统发育关系是多种生物学研究的基础，它被广泛运用于进化生物学、生物地理学、发育生物学、比较基因组学以及农作物育种中（Judd 等，1999）。随着上世纪80年代测序技术的迅速发展，分子系统学（molecular systematics）应运而生，为系统发育生物学提供了更加有力的工具（Soltis 等，2005）。现阶段，植物分子系统学应用的基因主要来自于叶绿体和线粒体，推动了分子系统学的发展（Jansen 等，2007；Qiu 等，2010）。但叶绿体和线粒体主要采取单亲遗传方式，不能完整记录物种的进化历史（Ness 等，2011）。另外，细胞器基因在进化过程中相对保守，物种之间的序列差异很小，不能提供足以解决物种间亲缘关系的信息位点（Lu 等，2010）。虽然经过20多年的研究，利用细胞器基因甚至是叶绿体全基因组重建的系统发育树仍无法全面解决物种间的系统发育关系（Jansen 等，2007；Moore 等，2007，2010）。

随着研究的不断深入，利用低拷贝核基因作为分子系统学的基因标记显得更加高效。这主要在于核基因属于双亲遗传，可以更好地反映物种的进化历史；而且，与细胞器基因相比，多数核基因可以提供更多的信息位点；另外，利用广泛保守的核基因便于将动物、植物及真菌等真核生物整合在一起，形成完整的生命之树。目前，核基因已经被广泛用于真菌和动物的系统发育重建并获得了更加详实的结论。例如 James 等（2006）利用6个编码蛋白的基因序列重建了约200种真菌之间的亲缘关系，结果表明，基于核基因得到的物种树可以得到更高的自展支持率。Regier 等（2010）分析了75种节肢动物的62个单拷贝核基因，将甲壳纲动物分为三个大的类群，结果表明核基因为节肢动物的分类提供了新的依据。另外，Moreau 等（2006）研究了蚂蚁中139个属的6个核基因编码区，分子钟（molecular clock）的分析结果表明，食草性昆虫的起源时间在白垩纪（cretaceous）到始新世（eocene）早期之间，从而推断蚂蚁类群的爆发和早期被子植物的起源有密切联系。

作为分子系统学研究工具的核基因必须为低拷贝基因，只有这样的基因才能减少旁系同源基因（paralog）的影响（Duarte 等，2010）。但是，对于绝大多数被子植物而言，它们在进化过程中曾发生过多次全基因组复制事件（whole-genome duplication，WGD）（Soltis 等，2009；Jiao 等，2011），使得植物的核基因组中出现了大量的重复基因，也加大了筛选低拷贝基因的难度。在实验方面，编码蛋白的核基因往往具有复杂的基因结构，这都对基因片段的扩增和测序提出了挑战。因此，筛选出一批能高效进行植物分子系统学研究的核基因就显得非常必要。

十字花科（Brassicaceae）在自然界中是一个庞大的自然类群，全科分为338个属，共有3 700多个种（Warwick 等，2006），并具有极高的经济和研究价值。拟南芥（*Arabidopsis thaliana*）和油菜（*Brassica rapa*），作为植物学研究的模式物种，是本科的代表物种。虽然经过数十年的发展，十字花科的系统关系却并没有得到彻底解决。十字花科物种进化速率较快，利用传统分类学方法鉴定的十字花科的族多为多系（polyphyletic group）或者并系（paraphyletic group）（Franzke 等，2009）。目前的研究结果中能够得到共识的有，醉碟花科植物（Cleomaceae）是整个十字花科的姐妹群，岩芥（*Aethionema*）是核心十字花科植物的姐妹群，核心十字花科分为3个大分支（Bailey 等，2006）。在十字花科的进化历史过程中曾经发生过3次全基因组复制事件，分别为At-α、At-β和At-γ，其中At-α和At-β都发生在在十字花科起源之前，而At-γ则更早（Soltis 等，2009）。Bailey 等（2006）利用核糖体基因转录间隔区（Internal Transcribed Spacers，ITS）分析了146个属，共计461个种，并将整个十字花科分为25个族。除了Camelineae以外，其余的族都属于单系类群（monophyletic group），但进化树主干上（backbone）的自展支持率却并不理想，这说明利用ITS序列未能解决十字花科各大类群之间的亲缘关系。Couvreur 等（2010）在结合前人的研究基础上，利用8个基因即 *nad4* intron 1、*ndhF*、*phyA*、*adh*、*chs*、ITS、*matK*、*trnLF* 分析了307个属的物种，继续细化了十字花科的分类，将其划分为至少33个族，但是分支内部的自展支持率仍然较低。综上所述，利用以上基因还无法彻底解决十字花科内部尤其是各个族之间的亲缘关系。

本研究中，我们选取了 5 个进化速率较快的核基因：*MCM5*（Mini-Chromosome Maintenance complex），*MLH1*（mutL Homolog；Mismatch Repair，MMR），*AT3G54630*，*AT3G26730* 和 *AT2G47990*。其中，*MCM5* 在所有真核生物中都存在，它与 *MCM2-4*，*6*，*7* 编码的蛋白亚基组成 DNA 解旋酶，在 DNA 复制过程中发挥关键作用（Maiorano 等，2006）；*MLH1* 在 DNA 修复过程中起到重要功能（Iyer 等，2006）；*AT3G54630*、*AT3G26730* 和 *AT2G47990* 是我们筛选出的进化速率较快的低拷贝核基因，功能未知。我们的研究结果表明，这些基因在基因组已测序的被子植物中大都只保留一个拷贝（未发表数据）。而且，*AT3G54630* 和 *AT2G47990* 基因中没有内含子。在本实验中，我们利用这 5 个基因来探究十字花科 17 个代表物种间的亲缘关系，从而验证它们在重建低等分类阶元亲缘关系中的适用性。

1　材料与方法

1.1　材料

本研究依据《中国植物志》第 33 卷的十字花科系统（周太炎等，1987）进行采样，共采集了 13 个物种作为研究对象，另外还从公共数据库中下载了拟南芥、琴叶拟南芥（*A. lyrata*）、白菜、盐芥（*Thellungiella salsuginea*）和番木瓜（*Carica papaya*）的基因序列，数据库网址见表 1。采集的植物样品放入 -80 ℃超低温冰箱中保存，样品名称、凭证标本及采集地点详情见表 2。

表 1　基因数据来源

Table 1　The websites of genetic data

物种 Species	网址 Website
拟南芥 *Arabidopsis thaliana*	www. arabidopsis. org
琴叶拟南芥 *A. lyrata*	www. phytozome. net
白菜 *Brassica rapa*	www. brassica. info
盐芥 *Thellungiella salsuginea*	thellungiella. org
番木瓜 *Carica papaya*	www. phytozome. net

1.2　RNA 提取及 cDNA 反转录

采用改进的 CTAB 法提取物种的总 RNA，将 2 g 植物嫩叶材料在液氮中研磨成粉，并加入到 4 mL CTAB 中（65 ℃预热），另外加入 40 μL 的 β-巯基乙醇防止 RNA 降解，后置于 65 ℃水浴锅 10 min。在溶液中加入 1/4 体积的氯仿：异戊醇（24：1）的萃取液，4 ℃和转速 12 000 r/min 的条件下离心 10 min，取上清并重复一次萃取步骤。将上清液移入新管子，加入 1/4 体积的 10 $mol \cdot L^{-1}$ 氯化锂，置于 -20 ℃冰箱中 6 ~ 8 h 沉淀总 RNA，离心收集沉淀，用 70% 的乙醇清洗两次，在通风橱内自然风干，溶于适量的双蒸水中，置于 -80 ℃保存。cDNA 合成使用 TaKaRa 公司生产的逆转录试剂盒（货号：D6210A），20 μM 反应体系中保证 RNA 的量约为 0.5 μg，反应条件为 42 ℃ 1 h，70 ℃ 15 min，最后将 cDNA 置于 -20 ℃冰箱中保存。

1.3　PCR 扩增及测序

利用拟南芥、琴叶拟南芥、白菜和盐芥的直系同源基因序列设计引物，*MCM5* 和 *MLH1* 的正反向引物上分别连接通用引物 T3 和 T7，引物长度为 40 ~ 45 bp；*AT3G54630*、*AT3G26730* 和 *AT3G47990* 的引物则无需连接通用引物，长度为 20 ~ 25 bp，具体序列见表 3。PCR

表 2　样品名称、凭证标本及采集地点

Table 2　Taxa，location and voucher included in this study

物种 Species	族 Tribes（周太炎系统）	采集地 Location	凭证标本 Voucher
播娘蒿 *Descurainia sophia*	大蒜芥族 Sisymbrieae	山东师范大学校园	M. Ding 1123
臭荠 *Coronopus didymus*	独行菜族 Lepidieae	上海植物园	M. Ding 1111
独行菜 *Lepidium apetalum*	独行菜族 Lepidieae	山东师范大学校园	M. Ding1122
诸葛菜 *Orychophragmus violaceus*	芸薹族 Brassiceae	上海植物园	M. Ding 1109
桂竹香 *Cheiranthus cheiri*	香花芥族 Hesperideae	复旦大学	M. Ding 1127
蔊菜 *Rorippa indica*	南芥族 Arabideae	上海植物园	M. Ding 1108
芥菜 *Brassica juncea*	芸薹族 Brassiceae	上海植物园	M. Ding 1106
荠菜 *Capsella bursa-pastoris*	独行菜族 Lepidieae	上海植物园	M. Ding 1107
碎米荠 *Cardamine hirsuta*	南芥族 Arabideae	上海植物园	M. Ding 1113
香雪球 *Lobularia maritima*	庭芥族 Alysseae	复旦大学江湾校区	M. Ding 1128
小花糖芥 *Erysimum cheiranthoides*	香花芥族 Hesperideae	山东师范大学校园	M. Ding 1124
萝卜 *Raphanus sativus*	芸薹族 Brassiceae	菜市场	M. Ding 1003
紫罗兰 *Matthiola incana*	紫罗兰族 Matthioleae	复旦大学	M. Ding 1133

表3 基因名、引物名及序列

Table 3 Sequence of primers used in this study

基因 Genes	引物 Primers	序列 Sequences
MCM5	T3-F1^{MCM5}	ATTAACCCTCACTAAAGGGAATHGAYGARTTTGAYAARATG
	T3-F2^{MCM5}	ATTAACCCTCACTAAAGGGA ATTCAYGARGCNATGGARCA
	T7-R1^{MCM5}	TAATACGACTCACTATAGGG ATDGAWTCRTTCATBCCCAT
MLH1	T3-F1^{MLH1}	ATTAACCCTCACTAAAGGGA GGRAATYTRTTYTAYAAYATG
	T3-F2^{MLH1}	ATTAACCCTCACTAAAGGGATCNTGYMGAAAGCATGGAGC
	T7-R1^{MLH1}	TAATACGACTCACTATAGGGTCCATRTCAGGIGTRTAYTGRTC
	T7-R1^{MLH1}	TAATACGACTCACTATAGGG ACRTGYTGDATIGWCCAYTC
AT3G54630	F1	GTCAARGASATCTCCGAAACCCTA
	F2	TAACTCCATGWMTTYCTTCG
	R1	TTCTCCTCCACAACYTTCTCCA
	R2	CTGAASCTCCACRCTCTTCTT
AT2G47990	F1	CCACCGTCTCCCTCTTCTCCTC
	R1	CTCAACGCCCTTCTCCTACTCTC
AT3G26730	F1	TAGCCATGGATCTTCCAAAGAG
	R1	CTGACAGAGAGATCAACATCTTG

反应使用TaKaRa公司生产的Ex Taq（货号：DRR006C）进行扩增。*MLH1* 和 *MCM5* 使用两轮PCR进行扩增，在第一轮PCR中，*MLH1* 选用T3-F1^{MLH1}/T7-R1^{MLH1}引物组合，*MCM5* 选用T3-F1^{MCM5}/T7-R1^{MCM5}引物组合，目的在于扩增出少量目的片段，在第二轮中，*MLH1* 选用T3-F2^{MLH1}/T7-R2^{MLH1}引物组合，*MCM5* 选用T3-F2^{MCM5}/T7-R1^{MCM5}引物组合，进一步富集目的片段。*AT3G54630*、*AT3G26730* 和 *AT2G47990* 的目的片段采用一次PCR扩增得到，其中 *AT3G26730* 和 *AT2G47990* 中只扩增一个基因片段；而 *AT3G54630* 扩增得到两个基因片段，引物组合分别为：F1/R1和F2/R2。反应程序为：94 ℃热启动3 min；然后进行25个循环，条件为94 ℃变性30 s，55 ℃退火30 s（每个循环温度降0.4 ℃），72 ℃延伸1 min（每个循环延伸时间增加2 s）；随后进行10个循环，条件为94 ℃变性30 s，45 ℃退火30 s，72 ℃延伸2 min（每个循环延伸时间增加3 s）。反应体系采用60 μL反应体系。PCR产物经琼脂糖凝胶电泳检测片段长度后，送上海赛英生物技术公司测序。

1.4 克隆实验

由于一些物种样本来自多棵植物，可能带有多个等位基因，PCR扩增产物不单一，直接测序结果表明为非单一序列。将这些PCR基因片段用来自Axygen公司（货号：L00359463）的胶回收试剂盒割胶回收。将纯化后的DNA片段连入TaKaRa公司的连接试剂盒中的18-T simple Vector载体（货号：D103A），16 ℃反应1 h后，将连接产物加入感受态细胞DH5α。扩大培养后倒入具有氨苄（AMP^+）抗性的固体培养基，放入37 ℃培养箱中倒置过夜培养。次日，挑取单菌落进行菌落PCR检验，并在具有氨苄抗性的液体培养基中扩大培养，选取电泳检测结果中具有正确条带的克隆进行测序，每个样品至少选取三个克隆进行测序。同一物种不同克隆之间的碱基差异低于1%，远少于不同物种之间的差异，而且同种的不同序列在单基因树中聚在一起（未发表结果），所以我们将相同物种不同克隆间的差异位点处理为兼并性位点（degenerate sites）。

1.5 外类群选择

本实验选取与十字花科亲缘关系较近的物种番木瓜作为外类群，番木瓜与十字花科均属于十字花目（Brassicales），由于番木瓜已经完成基因组测序，基因序列可以直接从Phytozome下载（表1）。

1.6 分子系统学常用分子标记序列下载

为了验证本实验中选取的5个核基因的高效性，我们还从GenBank中下载了17个物种（包括番木瓜）的3个常用分子标记的基因序列：*matK*（Maturase K），ITS（Internal transcribed spacer）和 *rbcL*（large subunit of ribulose-bisphosphate carboxylase）。其中，*matK* 和 *rbcL* 来源于植物叶绿体基因组，ITS是rDNA基因的间隔序列，本次研究中涉及到的ITS包括ITS1、ITS2和5.8SrRNA三部分。以上3个基因已经广泛用于十字花科系统关系研究中，各物种的基因序列号见表4。

1.7 数据分析

应用MEGA 5.0分别对5个核基因的DNA片段的序列进行序列比对，删除无法对齐的序列，并确定研究的序列是基因的开放读码框（Tamura 等，2011）。对5个

表 4　物种名、基因名以及 GenBank 序列号

Table 4　Accession numbers of three widely used phylogenetic markers

物种 Species	*matK*	ITS	*rbcL*	物种 Species	*matK*	ITS	*rbcL*
播娘蒿 *Descurainia sophia*	GQ424581.1	HQ896614.1	JN892038.1	小花糖芥 *Erysimum cheiranthoides*	HQ593289.1	DQ005989.1	HQ590092.1
臭荠 *Coronopus didymus*	——	AY254533.1	——	萝卜 *Raphanus sativus*	AB354262.1	GQ268079.1	GQ184382.1
独行菜 *Lepidium apetalum*	JF954300.1	JF976769.1	HM850119.1	紫罗兰 *Matthiola incana*	AF144361.1	DQ249848.1	HM850161.1
桂竹香 *Cheiranthus cheiri*	——	——	——	拟南芥 *Arabidopsis thaliana*	AF144378.1	AT3G41979.1	FN594842.1
焯菜 *Rorippa indica*	AF174530.1	AF128108.1	D88907.1	琴叶拟南芥 *A. lyrata*	AF144342.1	GQ303493.1	HM849785.1
芥菜 *Brassica juncea*	AB354274.1	FJ980406.1	AY167979.1	白菜 *Brassica rapa*	AB354276.1	JN564039.1	AY167976.1
荠菜 *Capsella bursa-pastoris*	AP009371.1	AF531561.1	AP009371.1	诸葛菜 *Orychophragmus violaceus*	EU306557.1	GQ261952.1	——
碎米荠 *Cardamine hirsuta*	HM850748.1	DQ268385.1	HM849847.1	盐芥 *Thellungiella salsuginea*	——	AF531626.1	——
香雪球 *Lobularia maritima*	AP009375.1	AY254530.1	AP009375.1	番木瓜 *Carica papaya*	EU431223.1	AY461547.1	EU431223.1

基因进行联合基因分析,利用 SeaView 将 5 个基因序列拼接成一个联合矩阵,构建系统发育树。利用 PhyML 3.0 进行最大似然法分析（maximum likelihood，ML）（Gouy 等，2010），在运用 ML 分析时，先利用 Modeltest 3.7 选择适合的分子进化模型，结果表明最合适的分子进化模型是 GTR + I + Γ（Posada 和 Crandall，1998）。利用 PAUP 4b10 进行最大简约性分析（maximum parsimony，MP），联合矩阵中的空位作为缺失处理，并对所有性状状态做等权无序处理，采用启发式搜索（heuristic search）1 000 次抽样，TBR 枝长交换，每步保留 10 棵树，获取严格一致树（Swofford，2002）。利用 MrBayes 3.1.2 进行贝叶斯分析，分析采用 MCMC（Markov Chain Monte Carlo process）算法（Ronqueist 和 Huelsenbeck，2003），根据 Modeltest 3.7 的结果设置相关参数，以随机树作为起始树，共运行 500 000 代，每 100 代进行 1 次抽样。摒弃前 25% 的老化样本（burn-in samples），用剩余样本构建一致树并计算各分支的后验概率（posterior probability，PP）。*matK*、ITS 和 *rbcL* 三个基因也采用相同的方法进行分析。

2　结果

2.1　所选核基因易于扩增

在采集的 13 个植物样品中均得到了 *MCM5* 和 *MLH1* 的目的条带，扩增效率为 100%；*AT3G54630* 分成前后两段进行扩增，其中 F1/R1 引物组合得到了所有 13 个物种的扩增条带，F2/R2 引物组合得到了 11 个样品的扩增条带，扩增效率为 92.3%（图 1）；另外，我们分别得到了 10 个物种的 *AT2G47990* 和 12 个物种的 *AT3G26730* 基因片段，扩增效率分别是 76.9% 和 92.3%。没有得到的基因序列均处理为数据缺失。5 个基因的测序结果，在 TAIR（www.arabidopsis.org）网站上进行 BLAST 搜索后表明我们得到的序列均是相应基因的同源基因。

2.2　核基因可以提供更多的信息位点

通过单基因之间的比较，我们发现核基因提供的信息位点数目远远多于 3 个常用的分子标记基因（表 5）。在 3 个已用于研究的基因中，ITS 提供了 158 个信息位点，占到序列总长度的 25.4%，*matK* 和 *rbcL* 提供的信息位点数占到各自序列长度的 6.5% 和 3%，可以看出，后两个基因能提供的信息位点数非常有限。本次研究所涉及的 5 个核基因中，*AT3G54630* 总长度为 1 227 个碱基，其中 446 个位点为信息位点，占到序列总长度的 36.3%；*AT3G26730* 总长度为 842 个碱基，提供了 179 个信息位点，占序列总长度的 21.3%；*AT2G47990* 总长度为 722 个碱基，提供了 219 个信息位点，占序列总长度的 30.3%；功能已知的 *MCM5* 和 *MLH1* 的信息位点数也分别占到各自序列长度的 19.7% 和 12.7%，都明显高于 *matK* 和 *rbcL*。

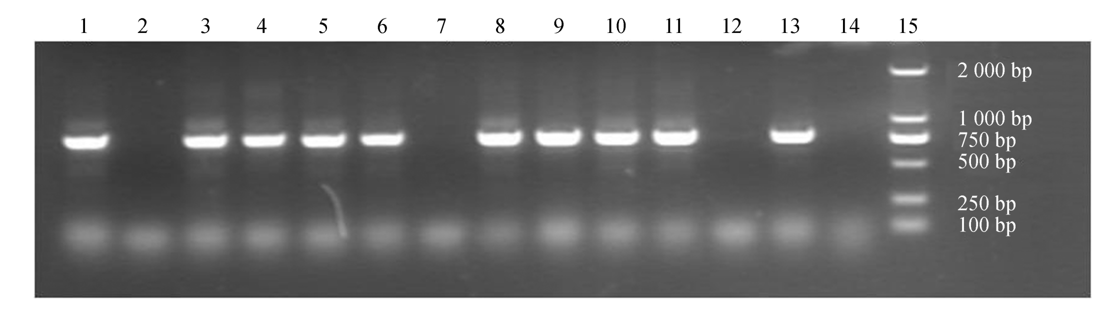


图1 十三个十字花科物种 *AT3G54630* 基因 PCR 片段电泳图，扩增引物为 F2/R2

1. 荠菜；2. 萝卜；3. 臭荠；4. 芥菜；5. 蔊菜；6. 二月兰；7. 香雪球；8. 碎米荠；9. 紫罗兰；10. 独行菜；11. 播娘蒿；12. 桂竹香；13. 小花糖芥；14. 空白对照；15. DL2000

Fig. 1 Electrophoresis of PCR products of *AT3G54630* that were amplified with F2/R2 from 13 Brassicaceae species

1. *Capsella bursa-pastoris*; 2. *Raphanus sativus*; 3. *Raphanus sativus*; 4. *Brassica juncea*; 5. *Rorippa indica*; 6. *Orychophragmus violaceus*; 7. *Lobularia maritima*; 8. *Cardamine hirsuta*; 9. *Matthiola incana*; 10. *Lepidium apetalum*; 11. *Descurainia sophia*; 12. *Cheiranthus cheiri*; 13. *Erysimum cheiranthoides*; 14. blank; 15. DL2000

表5 8个基因的系统学信息

Table 5 The systematic information of 8 genes

基因 Gene	序列长度 Gene length	变异位点数 Variable characters (VC)	信息位点数 Parsimony-informative characters (PIC)	信息位点数百分比 Ratio of PIC/%
AT3G26730	842	309	179	21.3
AT2G47990	722	354	219	30.3
AT3G54630	1227	724	446	36.3
MCM5	704	270	139	19.7
MLH1	1695	444	213	12.7
ITS	621	256	158	25.4
matK	1527	392	100	6.5
rbcL	1447	192	42	3

5个核基因序列经比对后联合成一个矩阵，总长度为4 585 bp，其中2 304个位点为变异位点，系统发育信息位点有1 233个，占整个序列的27%。应用MEGA 5.0计算5个核基因的碱基替换率均在0.2~0.3之间。作为对照组的3个常用分子标记基因，比对后联合为一个矩阵，总长度为3 624 bp，其中901个位点为变异位点，系统发育信息位点有314个，只占整个序列的8.7%，这3个常用分子标记的碱基替换率均在0.15以下。从联合基因分析上看，本文介绍的5个核基因提供了更多的信息位点。

2.3 基于核基因得到了支持率更高的物种关系

为了比较不同基因用于系统发育的效果，我们重建了多个单基因和联合基因系统发育树。从这些系统发育树上看，利用核基因 *AT3G54630* 建立的系统发育树不但好于利用3个常用分子标记基因（*matK*，*rbcL* 和 ITS）建立的单基因树，而且也好于基于3个常用分子标记建立的联合基因树。以 *AT3G54630* 和 *matK* 的MP树为例，利用 *AT3G54630* 建立的MP树中（图2），一致性指数（CI）为0.662，保留性指数（RI）为0.587，调整后一致性指数（RC）为0.389。紫罗兰位于十字花科的最基部，利用三种分析方法得到的支持率分别为100%（MP），78%（ML）和1.0（Bayesian）。除去诸葛菜和盐芥，其余14个十字花科植物分为两个分支，分别为 Clade I 和 Clade II。在 Clade I 中，拟南芥与琴叶拟南芥，荠菜与香雪球，小花糖芥与桂竹香，碎米荠与蔊菜，独行菜与臭荠分别组成姐妹群，均得到了最高支持，诸葛菜与 Clade I 组成姐妹群后再与盐芥组成姐妹群。在 Clade II 中，白菜与芥菜组成姐妹群，两者形成的分支再与萝卜组成姐妹群，3种分析方法均得到了最高的支持率。

利用 *matK* 建立的MP树中（图3），CI、RI

和 RC 分别为 0.873、0.703 和 0.614。香雪球位于十字花科的最基部，但只有 MP 法得到了 46% 的自展支持率，其余两种分析方法并不支持此结论。MP 分析方法中，紫罗兰、Clade I 和 Clade II 三支形成梳子状结构。在 Clade I 中，物种之间形成梳子状结构，亲缘关系无法解决。在 Clade II 中，白菜与芥菜组成姐妹群，得到了最高的支持率，它们再与萝卜组成姐妹群，得到的支持率

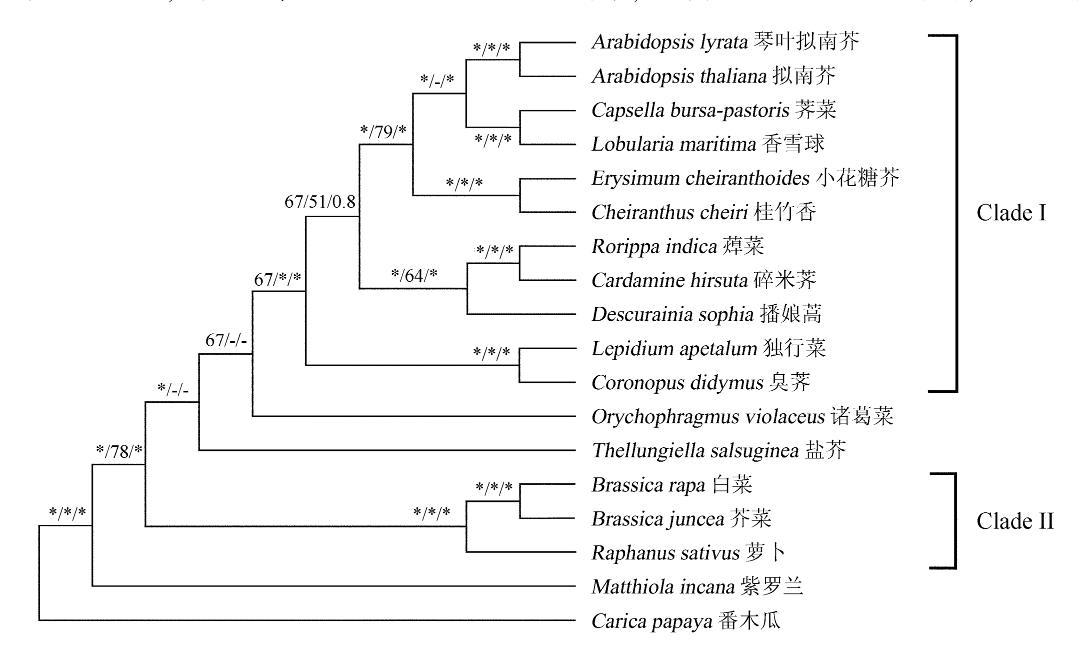


图 2　利用 *AT3G54630* 得到的 17 个物种的系统发育树，分支上的数字分别表示 MP/ML/Bayesian 分析得到的支持率及后验概率。星号代表自展支持率和后验概率分别为 100% 和 1.0

Fig. 2　MP tree based on *AT3G54630* for 17 Brassicaceae species

The numbers on each node are support values of MP, ML and Bayesian analyses, respectively. Asteriks stand for bootstrap of 100% or Posterior Probability of 1.0

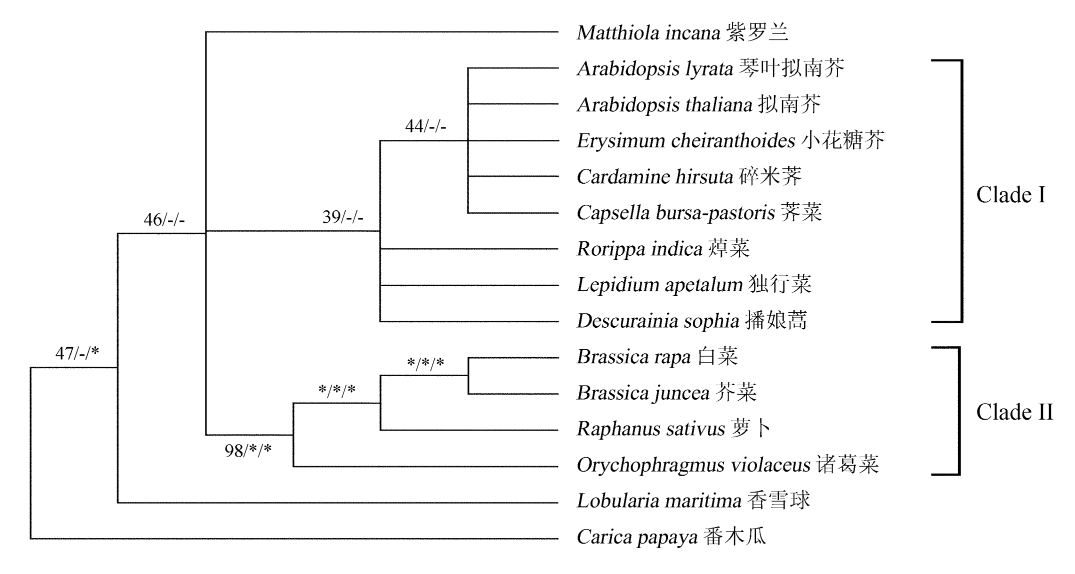


图 3　利用 *matK* 得到的 14 个物种的系统发育树，分支上的数字分别表示 MP/ML/Bayesian 分析得到的支持率及后验概率。星号代表自展支持率和后验概率分别为 100% 和 1.0

Fig. 3　MP tree based on *matK* for 14 Brassicaceae species

The numbers on each node are support values of MP, ML and Bayesian analyses, respectively. Asteriks stand for bootstrap of 100% or Posterior Probability of 1.0

也是最高，以上物种再与诸葛菜组成姐妹群，得到的支持率为 98% （MP），100% （ML）和 1.0（Bayesian）。由此可见，相对于 *AT3G54630*，*matK* 不能提供较好的分辨率。

基于 *matK*、*rbcL* 和 ITS 3 个基因的联合矩阵建立的 MP 树中（图 4），CI、RI 和 RC 分别为 0.730、0.510 和 0.372。在 Clade I 中，荠菜、拟南芥与琴叶拟南芥形成梳子状结构，小花糖芥与以上梳子结构形成姐妹群，得到的支持率分别为 83%（MP），54%（ML）和 1.0（Bayesian），碎米荠与蔊菜组成姐妹群，得到的支持率为 100%（MP），98%（ML）和 1.0（Bayesian），独行菜、播娘蒿和以上两大分支组成梳子状结构。在 Clade II 中，白菜与芥菜组成姐妹群，两者形成的分支再与萝卜组成姐妹群，诸葛菜再与以上物种形成姐妹群，支持率与 *AT3G54630* 的进化树相比都有所降低。臭荠和盐芥形成姐妹群，MP 法的支持率和后验概率分别为 100% 和 1.0，它们再与 Clade II 中的其他物种组成姐妹群。紫罗兰与 Clade I 和 Clade II 组成姐妹群后，再与香雪球组成姐妹群，只有 MP 法得到了 83% 的支持率。因此在本进化树中，利用 3 种分析方法得到的支持率并不高。可见，利用核基因 *AT3G54630* 得到的系统树不但比基于 *matK* 序列和 3 个基因建立的联合基因树得到更高的支持率，而且物种之间的关系也和已知关系具有较高的一致性。

利用 5 个核基因建立的联合基因树可以良好地展示 17 个十字花科物种之间的系统关系（图 5）。利用 MP 和 Bayesian 分析方法重建的联合基因树，每个节点的支持率均为 100% 。整棵树分为两个大的分支，在 Clade I 中，拟南芥和琴叶拟南芥，小花糖芥和桂竹香分别组成姐妹群后，两支之间组成姐妹群，3 种分析方法均得到了最高的支持率，以上物种形成的分支与播娘蒿组成姐妹群，MP 树的支持率和 Bayesian 树的后验概率分别为 100% 和 1.0。碎米荠与蔊菜组成姐妹群，两者形成的分支又与荠菜组成姐妹群，3 种分析方法均得到了最高的支持；独行菜与臭荠组成姐妹群后位于 Clade I 的基部。在 Clade II 中，白菜与芥菜组成姐妹群后与萝卜组成姐妹群，诸葛菜和盐芥依次与以上 3 个物种组成姐妹群。MP 树的支持率和 Bayesian 树的后验概率均为 100% 和 1.0，利用 ML 法得到的支持率至少 79% 。另外紫罗兰位于整个十字花科植物的基部，而香雪球则为除紫罗兰以外的基部分支，得到的自展支持率和后验概率均分别为 100% 和 1.0。

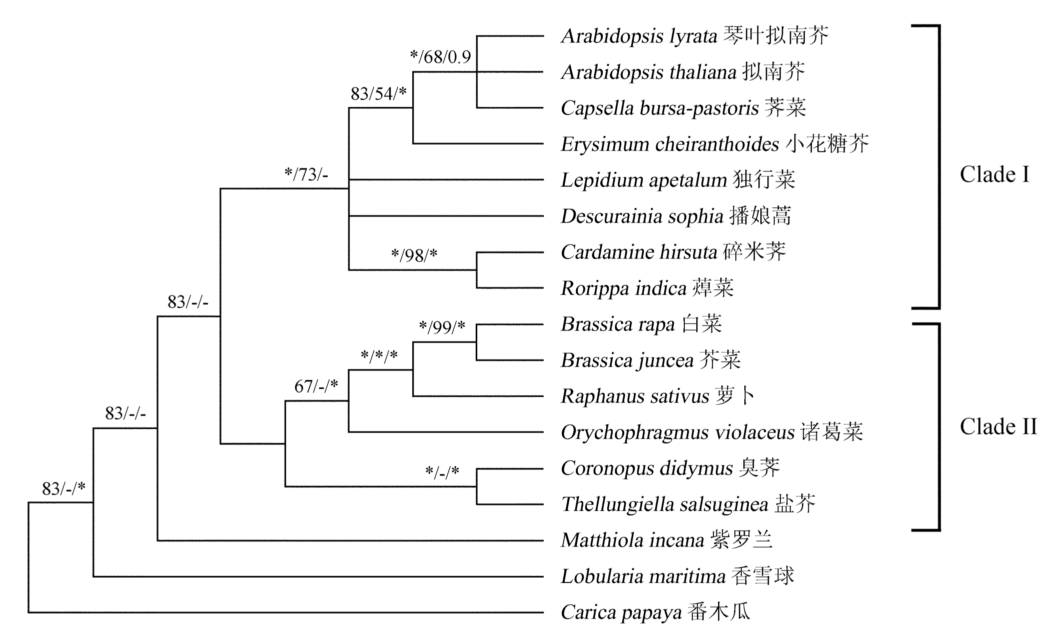


图 4 利用 3 个分子标记基因（*matK*，ITS 和 *rbcL*）联合矩阵得到的 17 个物种的系统发育树，分支上的数字分别表示 MP/ML/Bayesian 得到的支持率及后验概率。星号代表自展支持率和后验概率分别为 100% 和 1.0

Fig. 4 MP tree based on three concatenated molecular markers (*matK*, ITS, and *rbcL*) for 17 Brassicaceae species. The numbers on each node are support values of MP, ML and Bayesian, respectively. Asteriks stand for bootstrap of 100% and Posterior Probability of 1.0

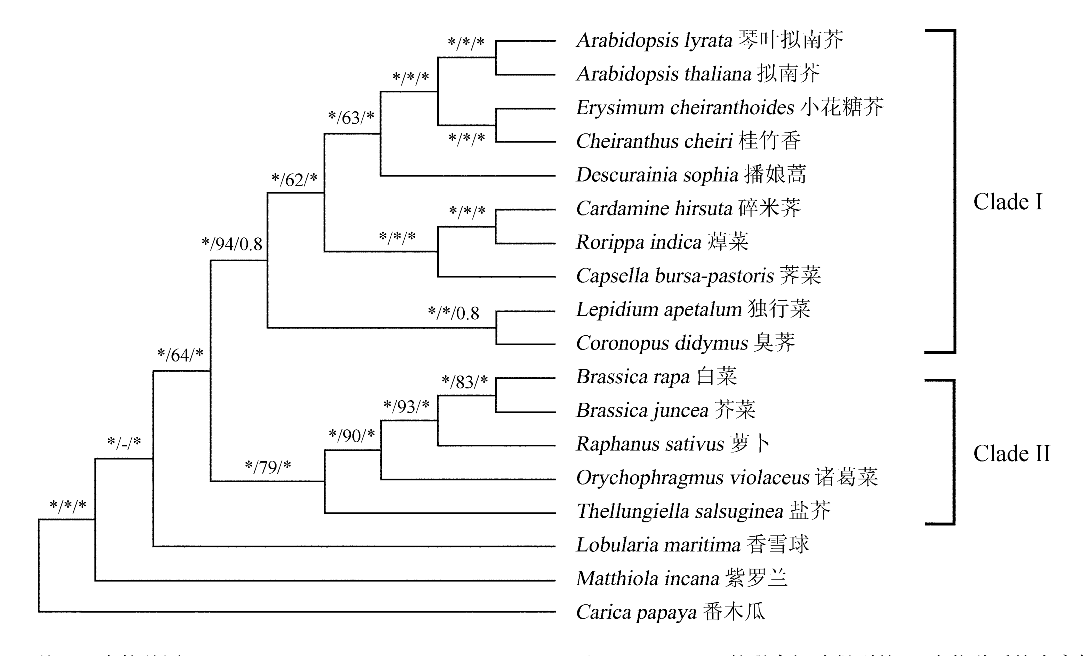


图 5　基于 5 个核基因（*MCM5*，*MLH1*，*AT3G54630*，*AT3G26730* 和 *AT2G47990*）的联合矩阵得到的 17 个物种系统发育树，分支上的数字分别表示 MP/ML/Bayesian 分析得到的支持率及后验概率。星号代表自展支持率和后验概率分别为 100% 和 1.0

Fig. 5　MP tree based on five concatenated nuclear genes (*MCM5*, *MLH1*, *AT3G54630*, *AT3G26730*, and *AT2G4799*) for 17 Brassicaceae species. The numbers on each node are support values of MP, ML and Bayesian

3　讨论

在被子植物分子系统学的发展过程中，现有研究大多利用细胞器基因重建物种间的亲缘关系。由于细胞器基因多为单亲遗传，而且物种间序列差异较小，使得被子植物分子系统还存在没有解决的问题（Chase 等，2009），在一些亲缘关系较近、进化速率较快的物种间，高度保守的细胞器基因的作用显得比较有限（Lu 等，2010）。另外，细胞器基因可能出现水平基因转移甚至从寄生植物中丢失，使得现有的分子标记不能有效地被运用到这些植物中，从而也不能被纳入生命之树（tree of life）（Keeling 和 Palmer，2008）。另一方面，核基因保留了双亲的遗传信息，较好地反映了物种的进化历史。与细胞器基因相比，大多数核基因在物种间差异较大，能提供更多的系统发育信息位点。目前少量的核基因已应用到分子系统学的研究中，以 ITS 为例，Bailey 等（2006）利用 ITS 分析了 146 个属的 461 个种，新的进化树细化了十字花科物种的系统关系，将整个十字花科分为 25 个族，但进化树主干的自展支持率仍然很低。ITS 序列为多拷贝的核糖体基因编码区之间的间隔序列，在某些类群中由于不完全地致同进化导致了拷贝间的差异，进而影响了系统发育重建的准确性；另外，ITS 序列长度也较短，提供的信息位点较少，可见仅仅依靠 ITS 并不能解决所有问题，尤其是低等分类阶层之间的关系。适合系统学研究的核基因应该具备 3 个条件：（1）基因应该具有保守序列，以便进行比对（alignment）；（2）基因的序列长度足够长，这样才能提供足够的信息位点；（3）用于研究的核基因应为直系同源基因。在本研究中，我们利用 5 个低拷贝核基因 *MCM5*、*MLH1*、*AT3G54630*、*AT3G26730* 和 *AT2G47990* 的编码区序列来研究 17 个十字花科植物之间的系统关系，以便验证这些核基因在系统分类上的适用性。

应用 MEGA 5.0 计算 5 个核基因的碱基替换率均在 0.2～0.3 之间，可见这些编码蛋白的核基因的进化速率较快，长达 4 585 bp 的矩阵共提供了 1 233 个信息位点，占整个序列的 27%。而 3 个常用的分子标记的碱基替换率均在 0.15 以下，变异速率较慢，在 3 624 bp 的序列中提供了 314 个信息位点，只占整个序列的 8.7%。另外，*AT3G54630* 和 *AT2G47990* 基因中没有内含子序列，扩增时可以直接用 DNA 进行扩增。因此，在物种采集时，无需在干冰中保存 RNA 样品，减少了材料采集时的困难，以 DNA 为模板的基

因扩增在实验操作上也更加快捷便利。

从得到的进化树上看出，利用 5 个核基因建立的系统进化树结构合理，物种之间的系统关系明确，用 MP、ML 及 Bayesian 3 种方法都得到了支持率很高的进化树，而利用 3 个常用分子标记建立的系统进化树却未能解决本研究中 17 个十字花科物种间的亲缘关系。例如，在利用常用分子标记基因得到的联合基因树中，拟南芥和琴叶拟南芥形成梳子状结构。另外，运用这 3 个常用分子标记中的单基因建立的系统进化树更是形成了大量的梳子状结构。可见，利用核基因建立的系统进化树具有更高的分辨率和支持率。

在利用 5 个核基因建立的系统进化树中，Clade I 中将过去确定亲缘关系较近的物种聚到了一支，并在自展支持率和后验概率上均得到了 100% 和 1.0 的支持率。另外，独行菜族中的物种臭荠，独行菜和荠属之间的关系有了新的诠释；在 Takhtajan 分类系统以及 Beilstein 利用叶绿体基因 *ndh*F 重建的十字花科系统关系中，臭荠、独行菜和荠并入独行菜族（Beilstein 等，2008）。但 Couvreur 通过研究 8 个基因表明，荠属与独行菜的关系较远，并不属于独行菜族，而独行菜与臭荠的关系较近；因此，Al-shehbaz 将臭荠（*Coronopus didymus*）改成 *Lepidium coronopus*，得到了 88% 的自展支持率（Al-Shehbaz 等，2002）。我们的结果表明臭荠和独行菜形成姐妹群，利用 3 种分析方法得到的支持率为 100%，100% 和 0.8，支持 Al-shehbaz 的说法。另外，在我们的研究中，小花糖芥和桂竹香是姐妹群，3 种方法得到的支持率均为最高。Yang 等（1998）利用 5s rRNA 序列的研究结果发现，碎米荠与焯菜在系统关系上很近，而本研究中，碎米荠和焯菜组成姐妹群，得到的支持率分别为 100%，100% 和 1.0，支持了 Yang 等的结论。拟南芥和琴叶拟南芥组成姐妹群关系，支持率分别为 100%，100% 和 1.0，说明利用 5 个核基因建立的联合基因树可以得到合理的系统关系。

在 Clade II 中，芸薹属中的两个物种，芥菜和白菜组成了姐妹群，自展支持率和后验概率分别为 100% 和 1.0，以上两个物种与萝卜组成姐妹群，利用 3 种方法得到的支持率分别为 100%，93% 和 1.0，诸葛菜又与以上 3 个物种形成了姐妹群，利用 3 种方法得到的支持率分别为 100%，90% 和 1.0。在传统分类中，以上 4 个物种都属于芸薹族，Bailey 等（2006）利用 ITS 研究了 461 个十字花科物种，Koch 等（2001）研究十字花科的线粒体基因 *nad4* intron1 同样也支持这样的分类关系，其中芥菜和白菜属于芸薹属，萝卜与芸薹属物种关系较近也得到了以上研究的支持。可见本研究中选取的核基因可以真实反映以上物种的亲缘关系。

在本次研究中，我们采集的物种只占中国十字花科植物的一小部分（约为 3%），并且采集地点多集中在华东地区，得到的系统关系并不能完整反映整个十字花科的系统关系。本次实验的目的只是以十字花科为例检验核基因在低等分类阶元系统分类上的适用性。分析结果表明，属于芸薹族的 4 个物种聚为一个大的分支，自展支持率和后验概率均为 100% 和 1.0，独行菜和臭荠，碎米荠和焯菜，小花糖芥和桂竹香，拟南芥和琴叶拟南芥之间均是姐妹群关系，自展支持率和后验概率也均为 100% 和 1.0。以上结果表明我们选取的编码蛋白的核基因在十字花科的系统分类中能得到与前人相似的结论。另外，在核基因建立的系统发育树中，无论是主干，还是属与属之间的关系都得到了极高的支持率，在 5 个核基因建立的联合基因树上，每个节点的自展支持率和后验概率均为 100% 和 1.0，可见本实验中所选取的 5 个核基因可以科学地揭示 17 个十字花科物种之间的亲缘关系。因此，我们可以利用这 5 个基因来研究其他低等分类阶元之间的系统发育关系、杂交引起的物种形成以及由于快速分化引起的近期物种爆发等问题。

致谢 复旦大学植物研究所郭长奎、尤辰江和崔洁在实验操作过程中给予指导；复旦大学生命科学院陆帆老师在标本鉴定过程中给予帮助；尹越在论文写作中提供宝贵意见。

〔参 考 文 献〕

周太炎，郭荣麟，蓝永珍等，1987. 中国植物志第 33 卷［M］. 北京：科学出版社

Al-Shehbaz IA，Mummenhoff K，Appel O，2002. *Cardaria*，*Coronopus*，and *Stroganowia* are united with *Lepidium*（Brassicaceae）［J］. *Novon*，**12**（1）：5—11

Bailey CD, Koch MA, Mayer M *et al.*, 2006. Toward a global phylogeny of the Brassicaceae [J]. *Molecular Biology and Evolution*, **23** (11): 2142—2160

Beilstein MA, Al-Shehbaz IA, Mathews S *et al.*, 2008. Brassicaceae phylogeny inferred from phytochrome A and *ndh*F sequence data: tribes and trichomes revisited [J]. *American Journal of Botany*, **95** (10): 1307—1327

Chase MW, Fay MF, Reveal JL *et al.*, 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III [J]. *Botanical Journal of the Linnean Society*, **161** (2): 105—121

Couvreur TLP, Franzker A, Al-Shehbaz IA *et al.*, 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae) [J]. *Molecular Biology and Evolution*, **27** (1): 55—71

Duarte JM, Wall PK, Edger PP *et al.*, 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels [J]. *BMC Evolutionary Biology*, **10** (1): 61—78

Franzke A, German D, Al-Shehbaz IA *et al.*, 2009. *Arabidopsis* family ties: molecular phylogeny and age estimates in the Brassicaceae [J]. *Taxon*, **58** (2): 425—437

Gouy M, Guindon S, Gascuel O *et al.*, 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building [J]. *Molecular Biology and Evolution*, **27** (2): 221—224

Iyer RR, Pluciennik A, Burdett V *et al.*, 2006. DNA mismatch repair: functions and mechanisms [J]. *Chemical Reviews*, **106** (2): 302—323

James TY, Kauff F, Schoch CL *et al.*, 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny [J]. *Nature*, **443** (7113): 818—822

Jansen RK, Cai ZQ, Raubeson LA *et al.*, 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns [J]. *Proceedings of the National Academy of Sciences*, *USA*, **104** (49): 19369—19374

Jiao YN, Wickett NJ, Ayyampalayam S *et al.*, 2011. Ancestral polyploidy in seed plants and angiosperms [J]. *Nature*, **473** (7345): 97—100

Judd WS, Campbell CS, Kellogg E *et al.*, 1999. *Plant Systematics: A Phylogenetic Approach. Sunderland*, *Massachusetts* [M]. Sunderland, Massachusetts: Sinauer Associates

Keeling PJ, Palmer JD, 2008. Horizontal gene transfer in eukaryotic evolution [J]. *Nature Reviews Genetics*, **9** (8): 605—618

Koch M, Haubold B, Mitchell-Olds T *et al.*, 2001. Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences [J]. *American Journal of Botany*, **88** (3): 534—544

Lu L, Fritsch PW, Cruz BC *et al.*, 2010. Reticulate evolution, cryptic species, and character convergence in the core East Asian clade of Gaultheria (Ericaceae) [J]. *Molecular Phylogenetics and Evolution*, **57** (1): 364—379

Maiorano D, Lutzmann M, Méchali M *et al.*, 2006. MCM proteins and DNA replication [J]. *Current Opinion in Cell Biology*, **18** (2): 130—136

Moore MJ, Bell CD, Soltis PS *et al.*, 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms [J]. *Proceedings of the National Academy of Sciences*, *USA*, **104** (49): 19363—19368

Moore MJ, Soltis PS, Bell CD *et al.*, 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots [J]. *Proceedings of the National Academy of Sciences*, *USA*, **107** (10): 4623—4628

Moreau CS, Bell CD, Vila R *et al.*, 2006. Phylogeny of the ants: diversification in the age of angiosperms [J]. *Science*, **312** (5570): 101—104

Ness RW, Graham SW, Barrett SCH, 2011. Reconciling gene and genome duplication events: using multiple nuclear gene families to infer the phylogeny of the aquatic plant family Pontederiaceae [J]. *Molecular Biology and Evolution*, **28** (11): 3009—3018

Posada D, Crandall KA, 1998. Modeltest: testing the model of DNA substitution [J]. *Bioinformatics*, **14** (9): 817—818

Qiu YL, Li L, Wang B *et al.*, 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes [J]. *Journal of Systematics and Evolution*, **48** (6): 391—425

Regier JC, Shultz JW, Zwick A *et al.*, 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences [J]. *Nature*, **463** (7284): 1079—1083

Ronqueist F, Huelsenbeck JP, 2002. MrBayes3: Bayesian phylogenetic inference under mixed models [J]. *Bioinformatics*, **19**: 1572—1574

Soltis DE, Soltis PS, Endress PK *et al.*, 2005. *Phylogeny and Evolution of Angiosperms* [M]. Sunderland, Massachusetts: Sinauer Associates

Soltis DE, Albert VA, Leebens-Mack J *et al.*, 2009. Polyploidy and angiosperm diversification [J]. *American Journal of Botany*, **96** (1): 336—348

Swofford DL, 2002. PAUP* 4. 0B10: *Phylogenetic Analysis Using Parsimony* (* *and other methods*), *Beta version* [M]. Sunderland: Sinauer Associates, Inc

Tamura K, Peterson D, Peterson N *et al.*, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods [J]. *Molecular Biology and Evolution*, **28** (10): 2731—2739

Warwick SI, Al-Shehbaz IA, Sauder CA *et al.*, 2006. Phylogenetic position of *Arabis arenicola* and generic limits of *Aphragmus* and *Eutrema* (Brassicaceae) based on sequences of nuclear ribosomal DNA [J]. *Molecular Biology and Evolution*, **84**: 269—281

Yang YW, Tseng PE, Tai PY *et al.*, 1998. Phylogenetic position of *Raphanus* in relation to *Brassica* species based on 5S rRNA spacer sequence data [J]. *Botanical Bulletin of Academia Sinica*, **39**: 153—160